Attorney Docket No. 944-3.43-1

UNITED STATES PATENT APPLICATION

of

Jari Mäkinen,

Hannu Mikkola,

Janne Vainio, and

Jani Rotola-Pukkila

for

AN IMPROVED SPECTRAL PARAMETER SUBSTITUTION FOR THE FRAME ERROR CONCEALMENT IN A SPEECH DECODER

AN IMPROVED SPECTRAL PARAMETER SUBSTITUTION FOR THE FRAME ERROR CONCEALMENT IN A SPEECH DECODER

CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims priority under 35 USC §119(e)(1) to provisional application Ser. No. 60/242,498 filed Oct. 23, 2000.

1995. FIELD OF THE INVENTION

The present invention relates to speech decoders, and more particularly to methods used to handle bad frames received by speech decoders.

BACKGROUND OF THE INVENTION

In digital cellular systems, a bit stream is said to be transmitted through a communication channel connecting a mobile station to a base station over the air interface. The bit stream is organized into frames, including speech frames. Whether or not an error occurs during transmission depends on prevailing channel conditions. A speech frame that is detected to contain errors is called simply a *bad frame*. According to the prior art, in case of a bad frame, speech parameters derived from past correct parameters (of non-erroneous speech frames) are substituted for the speech parameters of the bad frame. The aim of bad frame handling by making such a substitution is to conceal the corrupted speech parameters of the erroneous speech frame without causing a noticeable degrading of the speech quality.

Modern speech codecs operate by processing a speech signal in short segments, the above-mentioned frames. A typical frame

length of a speech codec is 20 ms, which corresponds to 160 speech samples, assuming an 8 kHz sampling frequency. In so-called wideband codecs, frame length can again be 20 ms, but can correspond to 320 speech samples, assuming a 16 kHz sampling frequency. A frame may be further divided into a number of subframes.

For every frame, an encoder determines a parametric representation of the input signal. The parameters are quantized and then transmitted through a communication channel in digital form. A decoder produces a synthesized speech signal based on the received parameters (see Fig. 1).

A typical set of extracted coding parameters includes spectral parameters (so called *linear predictive coding* parameters, or LPC parameters) used in short-term prediction, parameters used for long-term prediction of the signal (so called long-term prediction parameters or LTP parameters), various gain parameters, and finally, excitation parameters.

What is called linear predictive coding is a widely used and successful method for coding speech for transmission over a communication channel; it represents the frequency shaping attributes of the vocal tract. LPC parameterization characterizes the shape of the spectrum of a short segment of speech. The LPC parameters can be represented as either LSFs (Line Spectral Frequencies) or, equivalently, as ISPs (Immittance Spectral Pairs). ISPs are obtained by decomposing the inverse filter transfer function A(z) to a set of two transfer functions, one having even symmetry and the other having odd symmetry. The ISPs, also called Immittance Spectral Frequencies (ISFs), are the roots of these polynomials on the z-unit circle. Line Spectral Pairs (also called Line Spectral Frequencies) can be defined in the same way as Immittance Spectral Pairs; the difference between

these representations is the conversion algorithm, which transforms the LP filter coefficients into another LPC parameter representation (LSP or ISP).

Sometimes the condition of the communication channel through which the encoded speech parameters are transmitted is poor, causing errors in the bit stream, i.e. causing frame errors (and so causing bad frames). There are two kinds of frame errors: lost frames and corrupted frames. In a corrupted frame, only some of the parameters describing a particular speech segment (typically of 20 ms duration) are corrupted. In a lost frame type of frame error, a frame is either totally corrupted or is not received at all.

In a packet-based transmission system for communicating speech (a system in which a frame is usually conveyed as a single packet), such as is sometimes provided by an ordinary Internet connection, it is possible that a data packet (or frame) will never reach the intended receiver or that a data packet (or frame) will arrive so late that it cannot be used because of the real-time nature of spoken speech. Such a frame is called a lost frame. A corrupted frame in such a situation is a frame that does arrive (usually within a single packet) at the receiver but that contains some parameters that are in error, as indicated for example by a cyclic redundancy check (CRC). This is usually the situation in a circuit-switched connection, such as a connection in a system of the global system for mobile communication (GSM) connection, where the bit error rate (BER) in a corrupted frame is typically below 5%.

Thus, it can be seen that the optimal corrective response to an incidence of a bad frame is different for the two cases of bad frames (the corrupted frame and the lost frame). There are different responses because in case of corrupted frames, there

is unreliable information about the parameters, and in case of lost frames, no information is available.

According to the prior art, when an error is detected in a received speech frame, a substitution and muting procedure is begun; the speech parameters of the bad frame are replaced by attenuated or modified values from the previous good frame, although some of the least important parameters from the erroneous frame are used, e.g. the code excited linear prediction parameters (CELPs), or more simply the excitation parameters.

In some methods according to the prior art, a buffer is used (in the receiver) called the parameter history, where the last speech parameters received without error are stored. When a frame is received without error, the parameter history is updated and the speech parameters conveyed by the frame are used for decoding. When a bad frame is detected, via a CRC check or some other error detection method, a bad frame indicator (BFI) is set to true and parameter *concealment* (substitution for and muting of the corresponding bad frames) is then begun; the prior-art methods for parameter concealment use parameter history for concealing corrupted frames. As mentioned above, when a received frame is classified as a bad frame (BFI set to true), some speech parameters may be used from the bad frame; for example, in the example solution for corrupted frame substitution of a GSM AMR (adaptive multi-rate) speech codec given in ETSI (European Telecommunications Standards Institute) specification 06.91, the excitation vector from the channel is always used. When a speech frame is lost (including the situation where a frame arrives too late to be used, such as for example in some IP-based transmission systems), obviously no parameters are available from the lost frame to be used.

In some prior-art systems, the last good spectral parameters received are substituted for the spectral parameters of a bad frame, after being slightly shifted towards a constant predetermined mean. According to the GSM 06.91 ETSI specification, the concealment is done in LSF format, and is given by the following algorithm,

For i=0 to N-1:

LSF_q1(i)

$$=\alpha*past\_LSF\_q(i)+(1-\alpha)*mean\_LSF(i); \qquad (eq.\ 1.0)$$

LSF_q2(i)=LSF_q1(i);

where α = 0.95 and N is the order of the linear predictive (LP) filter being used. The quantity LSF_q1 is the quantized LSF vector of the second subframe, and the quantity LSF_q2 is the quantized LSF vector of the fourth subframe. The LSF vectors of the first and third subframes are interpolated from these two vectors. (The LSF vector for the first subframe in the frame n is interpolated from LSF vector of fourth subframe in the frame n-1, i.e. the previous frame). The quantity past_LSF_q is the quantity LSF_q2 from the previous frame. The quantity mean_LSF is a vector whose components are predetermined constants; the components do not depend on a decoded speech sequence. The quantity mean_LSF with constant components generates a constant speech spectrum.

Such prior-art systems always shift the spectrum coefficients towards constant quantities, here indicated as mean_LSF(i). The constant quantities are constructed by averaging over a long time period and over several successive talkers. Such systems therefore offer only a compromise solution, not a solution that is optimal for any particular speaker or situation; the tradeoff of the compromise is between

leaving annoying artifacts in the synthesized speech, and making the speech more natural in how it sounds (i.e. the quality of the synthesized speech).

What is needed is an improved spectral parameter substitution in case of a corrupted speech frame, possibly a substitution based on both an analysis of the speech parameter history and the erroneous frame. Suitable substitution for erroneous speech frames has a significant effect on the quality of the synthesized speech produced from the bit stream.

Summary Of The Invention

Accordingly, the present invention provides a method and corresponding apparatus for concealing the effects of frame errors in frames to be decoded by a decoder in providing synthesized speech, the frames being provided over a communication channel to the decoder, each frame providing parameters used by the decoder in synthesizing speech, the method including the steps of: determining whether a frame is a bad frame; and providing a substitution for the parameters of the bad frame based on an at least partly adaptive mean of the spectral parameters of a predetermined number of the most recently received good frames.

In a further aspect of the invention, the method also includes the step of determining whether the bad frame conveys stationary or non-stationary speech, and, in addition, the step of providing a substitution for the bad frame is performed in a way that depends on whether the bad frame conveys stationary or non-stationary speech. In a still further aspect of the invention, in case of a bad frame conveying stationary speech, the step of providing a substitution for the bad frame is performed using a mean of parameters of a predetermined number

of the most recently received good frames. In another still further aspect of the invention, in case of a bad frame conveying non-stationary speech, the step of providing a substitution for the bad frame is performed using at most a predetermined portion of a mean of parameters of a predetermined number of the most recently received good frames.

In another further aspect of the invention, the method also includes the step of determining whether the bad frame meets a predetermined criterion, and if so, using the bad frame instead of substituting for the bad frame. In a still further aspect of the invention with such a step, the predetermined criterion involves making one or more of four comparisons: an inter-frame comparison, an intra-frame comparison, a two-point comparison, and a single-point comparison.

From another perspective, the invention is a method for concealing the effects of frame errors in frames to be decoded by a decoder in providing synthesized speech, the frames being provided over a communication channel to the decoder, each frame providing parameters used by the decoder in synthesizing speech the method including the steps of: determining whether a frame is a bad frame; and providing a substitution for the parameters of the bad frame, a substitution in which past immittance spectral frequencies (ISFs) are shifted towards a partly adaptive mean given by:

$$ISF_q(i) = \alpha * past\_ISF_q(i) + (1-\alpha) * ISF_{mean}(i), \text{ for } i = 0..16,$$

where

α = 0.9,

$ISF_q(i)$ is the i^{th} component of the ISF vector for a current frame,

$past\_ISF_q(i)$ is the i^{th} component of the ISF vector from the previous frame,

$ISF_{mean}(i)$ is the i^{th} component of the vector that is a combination of the adaptive mean and the constant predetermined mean ISF vectors, and is calculated using the formula:

$$ISF_{mean}(i) = \beta * ISF_{const\_mean}(i) + (1-\beta) * ISF_{adaptive\_mean}(i), \text{ for } i = 0..16,$$

where β = 0.75, where $ISF_{adaptive\_mean}(i) = \frac{1}{3}\sum_{i=0}^{2} past\_ISF_q(i)$ and is updated whenever BFI =0 where BFI is a bad frame indicator, and where $ISF_{const\_mean}(i)$ is the i^{th} component of a vector formed from a long-time average of ISF vectors.

Brief Description of the Drawings

The above and other objects, features and advantages of the invention will become apparent from a consideration of the subsequent detailed description presented in connection with accompanying drawings, in which:

Fig. 1 is a block diagram of components of a system according to the prior art for transmitting or storing speech and audio signal;

Fig. 2 is a graph illustrating LSF coefficients [0 … 4kHz] of adjacent frames in a case of stationary speech, the Y-axis being frequency and the X-axis being frames;

Fig. 3. is a graph illustrating LSF coefficients [0 … 4kHz] of adjacent frames in case of non-stationary speech, the Y-axis being frequency and the X-axis being frames;

Fig. 4. is a graph illustrating absolute spectral deviation error in the prior-art method;

Fig. 5 is a graph illustrating absolute spectral deviation error in the present invention (showing that the present invention gives better substitution for spectral parameters than the prior-art method), where the highest bar in the graph (indicating the most probable residual) is approximately zero;

Fig. 6. is a schematic flow diagram illustrating how bits are classified according to some prior art when a bad frame is detected;

Fig. 7 is a flowchart of the overall method of the invention; and

Fig. 8 is a set of two graphs illustrating aspects of the criteria used to determine whether or not an LSF of a frame indicated as having errors is acceptable.

Best Mode For Carrying Out The Invention

According to the invention, when a bad frame is detected by a decoder after transmission of a speech signal through a communication channel (Fig. 1), the corrupted spectral parameters of the speech signal are concealed (by substituting other spectral parameters for them) based on an analysis of the spectral parameters recently communicated through the communication channel. It is important to effectively conceal corrupted spectral parameters of a bad frame not only because the corrupted spectral parameters may cause artifacts (audible sounds that are obviously not speech), but also because the subjective quality of subsequent error-free speech frames decreases (at least when linear predictive quantization is used).

An analysis according to the invention also makes use of the localized nature of the spectral impact of the spectral parameters, such as line spectral frequencies (LSFs). The spectral impact of LSFs is said to be localized in that if one LSF parameter is adversely altered by a quantization and coding process, the LP spectrum will change only near the frequency represented by the LSF parameter, leaving the rest of the spectrum unchanged.

The invention in general, for either a lost frame or a corrupt frame

According to the invention, an analyzer determines the spectral parameter concealment in case of a bad frame based on the history of previously received speech parameters. The analyzer determines the type of the decoded speech signal (i.e. whether it is stationary or non-stationary). The history of the speech parameters is used to classify the decoded speech signal (as stationary or not, and more specifically, as voiced or not); the history that is used can be derived mainly from the most recent values of LTP and spectral parameters.

The terms *stationary speech signal* and *voiced speech signal* are practically synonymous; a voiced speech sequence is usually a relatively stationary signal, while an unvoiced speech sequence is usually not. We use the terminology *stationary* and *non-stationary speech signals* here because that terminology is more precise.

A frame can be classified as voiced or unvoiced (and also stationary or non-stationary) according to the ratio of the power of the adaptive excitation to that of the total excitation, as indicated in the frame for the speech corresponding to the frame. (A frame contains parameters

according to which both adaptive and total excitation are constructed; after doing so, the total power can be calculated.)

If a speech sequence is stationary, the methods of the prior art by which corrupted spectral parameters are concealed, as indicated above, are not particularly effective. This is because stationary adjacent spectral parameters are changing slowly, so the previous good spectral values (not corrupted or lost spectral values) are usually good estimates for the next spectral coefficients, and more specifically, are better than the spectral parameters from the previous frame driven towards the constant mean, which the prior art would use in place of the bad spectral parameters (to conceal them). Fig. 2 illustrates, for a stationary speech signal (and more particularly a voiced speech signal), the characteristics of LSFs, as one example of spectral parameters; it illustrates LSF coefficients [0 … 4kHz] of adjacent frames of stationary speech, the Y-axis being frequency and the X-axis being frames, showing that the LSFs do change relatively slowly, from frame to frame, for stationary speech.

During stationary speech segments, concealment is performed according to the invention (for either lost or corrupted frames) using the following algorithm:

For i = 0 to N-1 (elements within a frame):

adaptive_mean_LSF_vector(i)

=(past_LSF_good(i)(0)+past_LSF_good(i)(1)+…+past_LSF_good(i)(K-1))/K;

LSF_q1(i)

$$=\alpha*past\_LSF\_good(i)(0)+(1-\alpha)*adaptive\_mean\_LSF(i); \qquad (2.1)$$

LSF_q2(i)=LSF_q1(i).

where α can be approximately 0.95, N is the order of LP filter, and K is the adaptation length. *LSF_q1(i)* is the quantized LSF

vector of the second subframe and *LSF_q2(i)* is the quantized LSF vector of the fourth subframe. The LSF vectors of the first and third subframes are interpolated from these two vectors. The quantity *past_LSF_good(i)(0)* is equal to the value of the quantity *LSF_q2(i-1)* from the previous good frame. The quantity *past_LSF_good(i)(n)* is a component of the vector of LSF parameters from the $n+1^{th}$ previous good frame (i.e. the good frame that precedes the present bad frame by n+1 frames). Finally, the quantity *adaptive_mean_LSF(i)* is the mean (arithmetic average) of the previous good LSF vectors (i.e. it is a component of a vector quantity, each component being a mean of the corresponding components of the previous good LSF vectors).

It has been demonstrated that the adaptive mean method of the invention improves the subjective quality of synthesized speech compared to the method of the prior art. The demonstration used simulations where speech is transmitted through an error-inducing communication channel. Each time a bad frame was detected, the spectral error was calculated. The spectral error was obtained by subtracting, from the original spectrum, the spectrum that was used for concealing during the bad frame. The absolute error is calculated by taking the absolute value from the spectral error. Figs. 4 and 5 show the histograms of absolute deviation error of LSFs for the prior art and for the invented method, respectively. The optimal error concealment has an error close to zero, i.e. when the error is close to zero, the spectral parameters used for concealing are very close to the original (corrupted or lost) spectral parameters. As can be seen from the histograms of Figs. 4 and 5, the adaptive mean method of the invention (Fig. 5) conceals

errors better than the prior-art method (Fig. 4) during stationary speech sequences.

As mentioned above, the spectral coefficients of non-stationary signals (or, less precisely, unvoiced signals) fluctuate between adjacent frames, as indicated in Fig. 3, which is a graph illustrating LSFs of adjacent frames in case of non-stationary speech, the Y-axis being frequency and the X-axis being frames. In such a case, the optimal concealment method is not the same as in the case of stationary speech signal. For non-stationary speech, the invention provides concealment for bad (corrupted or lost) non-stationary speech segments according to the following algorithm (the non-stationary algorithm):

For i = 0 to N-1:

partly_adaptive_mean_LSF(i)
*= β*mean_LSF(i) + (1-β)*adaptive_mean_LSF(i);* (2.3)

LSF_q1(i)
*= α*past_LSF_qood(i)(0)+(1-α)*partly_adaptive_mean_LSF(i);* (2.2)

LSF_q2(i)= LSF_q1(i);

where N is the order of the LP filter, where α is typically approximately 0.90, where *LSF_q1(i)* and *LSF_q2(i)* are two sets of LSF vectors for the current frame as in equation (2.1), where *past_LSF_q(i)* is *LSF_q2(i)* from the previous good frame, where *partly_adaptive_mean_LSF(i)* is a combination of the adaptive mean LSF vector and the average LSF vector, and where *adaptive_mean_LSF(i)* is the mean of the last K good LSF vectors (which is updated when BFI is not set), and where *mean_LSF(i)* is a constant average LSF and is generated during the design process of the codec being used to synthesize speech; it is an average LSF of some speech database. The parameter β is typically approximately 0.75, a value used to express the extent

to which the speech is stationary as opposed to non-stationary. (It is sometimes calculated based on the ratio of the long-term prediction excitation energy to the fixed codebook excitation energy, or more precisely, using the formula

$$\beta = \frac{1 + voiceFactor}{2}$$

where

$$voiceFactor = \frac{energy_{pitch} - energy_{innovation}}{energy_{pitch} + energy_{innovation}},$$

in which $energy_{pitch}$ is the energy of pitch excitation and $energy_{innovation}$ is the energy of the innovation code excitation. When most of the energy is in long-term prediction excitation, the speech being decoded is mostly stationary. When most of the energy is in the fixed codebook excitation, the speech is mostly non-stationary.)

For β = 1.0, equation (2.3) reduces to equation (1.0), which is the prior art. For β = 0.0, equation (2.3) reduces to the equation (2.1), which is used by the present invention for stationary segments. For complexity sensitive implementations (in applications where it is important to keep complexity to a reasonable level), β can be fixed to some compromise value, e.g. 0.75, for both stationary and non-stationary segments. Spectral parameter concealment specifically for lost frames.

In case of a *lost frame*, only the information of past spectral parameters is available. The substituted spectral parameters are calculated according to a criterion based on parameter histories of for example spectral and LTP (long-term prediction) values; LTP parameters include LTP gain and LTP lag value. LTP represents the correlation of a current frame to a

previous frame. For example, the criterion used to calculate the substituted spectral parameters can distinguish situations where the last good LSFs should be modified by an adaptive LSF mean or, as in the prior art, by a constant mean.

Alternative spectral parameter concealment specifically for corrupted frames

When a speech frame is *corrupted* (as opposed to lost), the concealment procedure of the invention can be further optimized. In such a case, the spectral parameters can be completely or partially correct when received in the speech decoder. For example, in a packet-based connection (as in an ordinary TCP/IP Internet connection), the corrupted frames concealment method is usually not possible because with TCP/IP type connections usually all bad frames are lost frames, but for other kinds of connections, such as in the circuit switched GSM or EDGE connections, the corrupted frames concealment method of the invention *can* be used. Thus, for packet-switched connections, the following alternative method cannot be used, but for circuit-switched connections, it *can* be used, since in such connections bad frames are at least sometimes (and in fact usually) only corrupted frames.

According to the specifications for GSM, a bad frame is detected when a BFI flag is set following a CRC check or other error detection mechanism used in the channel decoding process. Error detection mechanisms are used to detect errors in the subjectively most significant bits, i.e. those bits having the greatest effect on the quality of the synthesized speech. In some prior art methods, these most significant bits are not used when a frame is indicated to be a bad frame. However, a frame may have only a few bit errors (even one being enough to set the BFI flag), so the whole frame could be discarded even though

most of the bits are correct. A CRC check detects simply whether or not a frame has erroneous frames, but makes no estimate of the BER (bit error rate). Fig. 6 illustrates how bits are classified according to the prior art when a bad frame is detected. In Fig. 6, a single frame is shown being communicated, one bit at a time (from left to right), to a decoder over a communications channel with conditions such that some bits of the frame included in a CRC check are corrupted, and so the BFI is set to one.

As can be seen from Fig. 6, even when a received frame sometimes contains many correct bits (the BER in a frame usually being small when channel conditions are relatively good), the prior art does not use them. In contrast, the present invention tries to estimate if the received parameters are corrupted and if they are not, the invented method uses them.

Table 1 demonstrates the idea behind the corrupted frame concealment according to the invention in the example of an adaptive multi-rate (AMR) wideband (WB) decoder.

	C/I [dB]				
mode 12.65 (AMR-WB)	10	9	8	7	6
BER	3.72%	4.58%	5.56%	6.70%	7.98%
FER	0.30%	0.74%	1.62%	3.45%	7.16%
Correct spectral parameter indexes	84%	77%	68%	64%	60%
Totally corrcet spectrum	47%	38%	32%	27%	24%

Table 1. Percentage of correct spectral parameters in a corrupted speech frame.

In case of an AMR WB decoder, mode 12.65 kbit/s is a good choice to use when the channel carrier to interference ratio (C/I) is in the range from approximately 9 dB to 10 dB. From Table 1, it can be seen that in case of GSM channel conditions with a C/I in the range 9 to 10 dB using a GMSK (Gaussian Minimum-Shift Keying) modulation scheme, approximately 35-50% of received bad

frames have a totally correct spectrum. Also, approximately 75-85% of all bad frame spectral parameter coefficients are correct. Because of the localized nature of the spectral impact, as mentioned earlier, spectral parameter information can be used in the bad frames. Channel conditions with a C/I in the range 6-8 dB or less are so poor that the 12.65 kbit/s mode should not be used; instead, some other, lower mode should be used.

The basic idea of the present invention in the case of corrupted frames is that according to a criterion (described below), channel bits from a corrupt frame are used for decoding the corrupt frame. The criterion for spectral coefficients is based on the past values of the speech parameters of the signal being decoded. When a bad frame is detected, the received LSFs or other spectral parameters communicated over the channel are used if the criterion is met; in other words, if the received LSFs meet the criterion, they are used in decoding just as they would be if the frame were not a bad frame. Otherwise, i.e. if the LSFs from the channel do not meet the criterion, the spectrum for a bad frame is calculated according to the concealment method described above, using equations (2.1) or (2.2). The criterion for accepting the spectral parameters can be implemented by using for example a spectral distance calculation such as a calculation of the so-called Itakura-Saito spectral distance. (See, for example, page 329 of *Discrete-Time Processing of Speech Signals* by John R Deller Jr, John H.L. Hansen, and John G. Proakis,, published by IEEE Press, 2000.)

The criterion for accepting the spectral parameters from the channel should be very strict in the case of a stationary speech signal. As shown in Fig. 3, the spectral coefficients are very stable during a stationary sequence (by definition) so

that corrupted LSFs (or other speech parameters) of a stationary speech signal can usually be readily detected (since they would be distinguishable from uncorrupted LSFs on the basis that they would differ dramatically from the LSFs of uncorrupted adjacent frames). On the other hand, for a non-stationary speech signal, the criterion need not be so strict; the spectrum for a non-stationary speech signal is allowed to have a larger variation. For a non-stationary speech signal, the exactness of the correct spectral parameters is not strict in respect to audible artifacts, since for non-stationary speech (i.e. more or less unvoiced speech), no audible artifacts are likely regardless of whether or not the speech parameters are correct. In other words, even if bits of the spectral parameters are corrupted, they can still be acceptable according to the criterion, since spectral parameters for non-stationary speech with some corrupt bits will not usually generate any audible artifacts. According to the invention, the subjective quality of the synthesized speech is to be diminished as little as possible in case of corrupted frames by using all the available information about the received LSFs, and by selecting which LSFs to use according to the characteristics of the speech being conveyed.

Thus, although the invention includes a method for concealing corrupted frames, it also comprehends as an alternative using a criterion in case of a corrupted frame conveying non-stationary speech, which, if met, will cause the decoder to use the corrupted frame as is; in other words, even though the BFI is set, the frame will be used. The criterion is in essence a threshold used to distinguish between a corrupted frame that is useable and one that is not; the threshold is based on how much the spectral parameters of the corrupted frame differ from the spectral parameters of the most recently received good frames.

The use of possible corrupted spectral parameters is probably more sensitive to audible artifacts than use of other corrupted parameters, such as corrupted LTP lag values. For this reason, the criterion used to determine whether or not to use a possibly corrupt spectral parameter should be especially reliable. In some embodiments, it is advantageous to use as the criterion a maximum spectral distance (from a corresponding spectral parameter in a previous frame, beyond which the suspect spectral parameter is not to be used); in such an embodiment, the well-known Itakura-Saito distance calculation could be used to quantify the spectral distance to be compared with the threshold. Alternatively, fixed or adaptive statistics of spectral parameters could be used for determining whether or not to use possibly corrupted spectral parameters. Also other speech parameters, such as gain parameters, could be used for generating the criterion. (If the other speech parameters are not drastically different in the current frame, compared to the values in the most recent good frame, then the spectral parameters are probably okay to use, provided the received spectral parameters also meet the criteria. In other words, other parameters, such as LTP gain, can be used as an additional component to set proper criteria to determine whether or not to use the received spectral parameters. The history of the other speech parameters can be used for improved recognition of speech characteristic. For example, the history can be used to decide whether the decoded speech sequence has a stationary or non-stationary characteristic. When the properties of the decoded speech sequence are known, it is easier to detect possibly correct spectral parameters from the corrupted frame and it is easier to estimate what kind of spectral parameter values are expected to have been conveyed in a received corrupted frame.)

According to the invention in the preferred embodiment, and now referring to Fig. 8, the criterion for determining whether or not to use a spectral parameter for a corrupted frame is based on the notion of a spectral distance, as mentioned above. More specifically, to determine whether the criterion for accepting the LSF coefficients of a corrupted frame is met, a processor of the receiver executes an algorithm that checks how much the LSF coefficients have moved along the frequency axis compared to the LSF coefficients of the last good frame, which is stored in an LSF buffer, along with the LSF coefficients of some predetermined number of earlier, most recent frames.

The criterion according to the preferred embodiment involves making one or more of four comparisons: an inter-frame comparison, an intra-frame comparison, a two-point comparison, and a single-point comparison.

In the first comparison, the inter-frame comparison, the differences between LSF vector elements in adjacent frames of the corrupted frame are compared to the corresponding differences of previous frames. The differences are determined as follows:

$$d_n(i) = \left| L_{n-1}(i) - L_n(i) \right|, \qquad 1 \le i \le P-1,$$

where P is the number of spectral coefficients for a frame, *$L_n(i)$ is the i^{th} LSF element of corrupted frame, and $L_{n-1}(i)$ is the i^{th} LSF element of the frame before corrupted frame.* The LSF element, $L_n(i)$, of the corrupted frame is discarded if the difference, $d_n(i)$, is too high compared to $d_{n-1}(i)$, $d_{n-2}(i)$,..., $d_{n-k}(i)$, where k is the length of the LSF buffer.

The second comparison, the intra-frame comparison, is a comparison of difference between adjacent LSF vector elements in the same frame. The distance between the candidate i^{th} LSF

element, $L_n(i)$, of the n^{th} *frame and the* $(i-1)^{th}$ LSF element, $L_{n-1}(i)$, of the n^{th} *frame is determined as follows:*

$$e_n(i) = L_n(i-1) - L_n(i), \qquad 2 \le i \le P-1,$$

where P is the number of spectral coefficients and $e_n(i)$ *is the distance between LSF elements.* Distances are calculated between all LSF vector elements of the frame. One or another or both of the LSF elements $L_n(i)$ and $L_n(i-1)$ will be discarded if the difference, $e_n(i)$, is too large or too small compared to $e_{n-1}(i)$, $e_{n-2}(i)$,..., $e_{n-k}(i)$.

The third comparison, the two-point comparison, determines whether a crossover has occurred involving the candidate LSF element $L_n(i)$, i.e. whether an element $L_n(i-1)$ that is lower in order than the candidate element has a larger value than the candidate LSF element $L_n(i)$. A crossover indicates one or more highly corrupted LSF values. All crossing LSF elements are usually discarded.

The fourth comparison, the single-point comparison, compares the value of the candidate LSF vector element, $L_n(i)$ to a minimum LSF element, $L_{min}(i)$, and to a maximum LSF element, $L_{max}(i)$, both calculated from the LSF buffer, and discards the candidate LSF element if it lies outside the range bracketed by the minimum and maximum LSF elements.

If an LSF element of a corrupted frame is discarded (based on the above criterion or otherwise), then a new value for the LSF element is calculated according to the algorithm using equation (2.2).

Referring now to Fig. 7, a flowchart of the overall method of the invention is shown, indicating the different provisions

for stationary and non-stationary speech frames, and for corrupted as opposed to lost non-stationary speech frames.

Discussion

The invention can be applied in a speech decoder in either a mobile station or a mobile network element. It can also be applied to any speech decoder used in a system having an erroneous transmission channel.

Scope of the Invention

It is to be understood that the above-described arrangements are only illustrative of the application of the principles of the present invention. In particular, it should be understood that although the invention has been shown and described using line spectrum pairs for a concrete illustration, the invention also comprehends using other, equivalent parameters, such as immittance spectral pairs. Numerous modifications and alternative arrangements may be devised by those skilled in the art without departing from the spirit and scope of the present invention, and the appended claims are intended to cover such modifications and arrangements.